第39卷 第1期
2022年1月

干 旱 区 研 究
ARID ZONE RESEARCH

Vol. 39 No. 1
Jan. 2022

基于3种机器学习方法的农业干旱监测比较

王晓燕， 李 净， 邢立亭

（西北师范大学地理与环境科学学院，甘肃 兰州 730070）

摘 要： 频繁的旱灾对甘肃省经济、农业生产造成严重危害，因此利用先进方法建立准确可靠的农业干旱监测模型对该省防旱减灾十分重要。本文基于随机森林（RF）、BP神经网络和支持向量机（SVM）3种机器学习方法，利用甘肃省2002—2019年4—10月多源遥感数据得到的植被状态指数（VCI）、温度状态指数（TCI）、植被供水指数（VSWI）、降水状态指数（PCI）以及DEM、土壤有效含水量（AWC）和气候类型作为自变量，气象站点以3个月时间尺度的标准化降水蒸发指数（SPEI_3）为因变量，构建3种不同的农业干旱监测模型，分析比较出适用于监测甘肃省农业干旱的最佳模型，同时进一步探究了机器学习方法构建的模型在不同环境下的适用性。结果表明：构建的3种机器学习模型中，随机森林模型的R^2平均值高（0.86）且误差小（RMSE为0.40，MAE为0.31），农业干旱的监测效果要优于BP神经网络和支持向量机模型；干燥和湿润两种环境下分别构建的3种机器学习模型在湿润环境中监测能力表现均更优异（R^2>0.82），而随机森林模型在两种环境中监测干旱的表现比其他两种模型强。研究结果为甘肃省的农业干旱监测与评估提供了新的科学方法，对农业干旱研究具有重要意义。

关键词： 农业干旱；机器学习；SPEI；MODIS

干旱是出现频率高、持续时间长、波及范围广的气象灾害之一[1]。农业干旱是由于低于正常降水或高于平均蒸发和蒸腾作用而导致的土壤水分亏缺[2]。频繁的农业干旱导致经济损失严重，因此准确且实时或接近实时的农业干旱监测是必不可少的。

干旱指数是监测和分析农业干旱的重要方法。近年来，学者们建立了各种干旱指数来量化干旱特征，特别是严重程度和空间范围[3]。根据数据源不同，通常分为基于气象站点数据监测的干旱指数和基于遥感数据监测的干旱指数两大类。常用的基于气象站点数据的干旱指数包括帕默尔旱度指数（PDSI）[4]、作物干旱识别指数[5]、综合指数（CI）[6]、K指数[7]、标准化降水指数（SPI）[8]和标准化降水蒸发指数（SPEI）[9-10]等。SPEI由于同时考虑了降水量和温度并且能够监测不同地区不同类型的干旱，被广泛应用。虽然基于气象站点数据的干旱指数能够准确且有效的监测气象站及其周边干旱的严重程度，但在可用气象站点有限的条件下，遥感数据覆盖面积广、空间分辨率高、时效性强等特点，基于遥感数据的干旱指数在大面积干旱时空模式的监测中更为可靠。当前基于遥感数据监测的干旱指数主要有归一化植被指数（NDVI）[11]、植被状态指数（VCI）[12]、温度状态指数（TCI）[13]、归一化多波段干旱指数（NMDI）[14]、归一化差值水分指数（NDWI）[15]以及植被供水指数（VSWI）[16]等。最初多是单一因素的遥感指数被用于监测和分析干旱，如NDVI、VCI等，但农业干旱过程复杂多变，影响因素众多，单一因素的干旱指数往往难以反映干旱的多类型和多尺度特征[17]。因此利用先进方法集成多源数据构建综合的干旱监测模型或指数逐渐成为发展新趋势。

综合多源数据的干旱监测模型或指数从构建方法角度介绍，有权重组合、多变量联合分布及机器学习方法。Zhang等[18]对土壤水分条件指数（SMCI）、TCI、PCI进行线性加权构建微波集成干旱指数（MIDI），结果表明MIDI与实测值在空间分布上有很好的一致。杜瑞麒等[19]通过联合分布函数

收稿日期： 2021-03-12； **修订日期：** 2021-06-02
基金项目： 中科院“西部之光”人才项目；国家自然科学基金项目（41861013，42071089）
作者简介： 王晓燕（1997-），女，硕士研究生，主要从事干旱研究. E-mail: 1748417278@qq.com
通讯作者： 李净. E-mail: li_jinger@163.com

构建出反映土壤水分和降水信息的综合干旱指数，并较好的分析了吉林省的干旱时空特征。虽然权重组合与联合分布方法在干旱监测中均有较好的表现，但均存在一定程度的缺陷。权重组合方法建立的综合干旱指数需要先假设变量间的线性关系然后根据专家经验判断或相关分析等方法来确定多种单一干旱指标的权重，而农业干旱的复杂决定了影响因子间不一定都是线性关系，同时权重的分配也存在误差；利用联合分布方法构建的综合干旱指数虽然既保留了单一指标本身的边际分布，又描述了指标之间复杂的依存关系，但在单一指标数量较多的情况下，变量间的关系变得复杂时，会导致建模变得困难。近年机器学习方法逐渐成为干旱研究领域的新秀，因为其不仅能够处理干旱影响因子间复杂的非线性问题且能够简单明了的集成多源数据，科学高效的建立综合干旱监测模型，合理的解决了上述问题。但是不同机器学习方法构建的模型在农业干旱监测上具有区域差异性，而且甘肃省气候类型复杂，干旱频发。为了能在甘肃省建立更加全面精准的综合农业干旱监测模型，本文采用随机森林、BP神经网络、支持向量机3种机器学习方法，考虑多种干旱因子，结合气象数据和多源遥感数据，比较分析构建的3种综合干旱监测模型在甘肃省的适用性，同时探究模型在不同环境下的表现并分析不同致旱因子的相对重要性，为农业干旱监测研究提供了新方法以及科学参考。

1 数据与方法

1.1 研究区概况

甘肃省位于中国西北部，地理位置为32°11′~42°57′N，92°13′~108°46′E。甘肃是典型的温带大陆性气候，降水少蒸发大。在甘肃几乎每年都发生农业干旱，每年平均受灾面积达到82.68×10^4 km²，并导致5×10^8~10×10^8 kg的粮食减产[20]。

1.2 数据处理

本文从气象、土壤、植被方面进行因子选择，选取PCI和TCI为气象因子，VSWI作为土壤因子，VCI是植被因子。考虑到研究区内不同地形水分、温度、植被覆盖度等因素在空间上存在差异性，土壤的生产力和抗旱性在空间上有高度变异性以及研究区内气候类型复杂等原因，选择数字高程模型DEM、土壤有效含水量以及中国气候区划为模型的辅助因子。

数据选择的时间范围为2002—2019年，时间分辨率为月（4—10月），空间分辨率为1 km。其中，MODIS数据有MOD11A2产品中的地表温度（LST），时间分辨率为8 d，空间分辨率为1 km，月数据为4幅影像数据求平均；MOD13A2产品中的归一化植被指数和增强型植被指数（EVI），时间分辨率为16 d，空间分辨率为1 km。降雨数据选择TRMM3B43，时间分辨率为月，空间分辨率为0.25°。以上MODIS和TRMM数据均来自于NASA（https://ladsweb.modaps.eosdis.nasa.gov/），在数据处理过程中，对于TRMM数据首先将降水速率转换为月降水总量数据，然后统一对NDVI、LST、EVI以及TRMM数据进行裁剪、投影转换、采用最邻近法进行重采样使空间分辨率为1 km，最后通过以上数据计算出VCI、TCI、PCI、VSWI指数，具体计算过程如表1。

DEM数字高程、中国气候区划和中国土壤砂、黏粒含量数据以及用于制图的土地覆盖类型数据均来源于中国科学院资源环境科学数据中心（https://www.resdc.cn/），空间分辨率均为1 km。土壤有效含水量利用Gupta等[21]提出的土壤砂、黏粒含量的经验线性拟合模型估算。对全国DEM、气候区划以及土壤有效含水量进行裁剪、投影转换后可直接使用。

气象数据来源于中国气象数据网（http://data.

表1 干旱指数计算方法

Tab. 1 Remote sensing drought index calculation formula

干旱指数	公式	文献来源
VCI植被状态指数	$(NDVI_i - NDVI_{min})/(NDVI_{max} - NDVI_{min})$	[22]
TCI温度状态指数	$(LST_{max} - LST_i)/(LST_{max} - LST_{min})$	[22]
PCI降水指数	$(TRMM_i - TRMM_{min})/(TRMM_{max} - TRMM_{min})$	[22]
VSWI植被供水指数	EVI/LST	[23]

注：i为第i月的NDVI、LST、TRMM值（i=4，5，6，…，10），max与min分别为2002—2019年第i月的最大值与最小值。

cma.cn/)，气象站点分布如图1所示，根据研究区气象数据的缺失和站点分布在耕地的情况，最终选择25个气象站点。站点数据主要包括1970—2019年的月平均气温和月降水量，用于计算1个月、3个月、6个月时间尺度的SPEI。

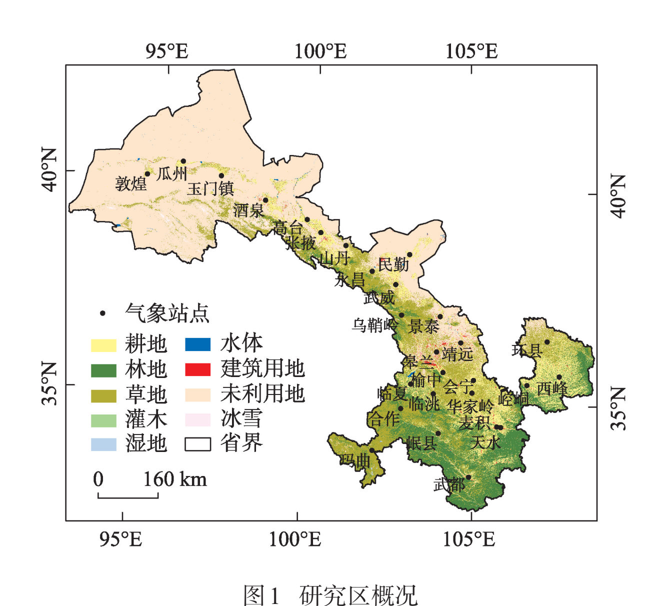


图1 研究区概况

Fig. 1 Overview of the study area

1.3 研究方法

1.3.1 随机森林　随机森林的核心思想是回归和分类，是使用去相关树有效地减少预测变量方差的一种方法[24]。随机森林方法的具体步骤为：首先从训练数据中，通过随机提取训练数据集中的部分数据，比如20%的数据集，并据此形成决策树；然后将提取的20%的数据集放回训练集中，并再次随机提取20%的数据集，产生第二个决策树。不断重复第二步，建立多个决策树并以此组成随机森林；最终预测结果由所有决策树结果的平均值决定。

1.3.2 BP神经网络　BP神经网络是一种由输入层、隐层和输出层组成，并由误差逆传播方法训练的多层前馈网络[25]。主要包括前向传播和误差的反向传播。前向传播指在计算误差输出时，输入信号通过隐含层作用于输出节点，经过非线性变换，产生输出信号，如果实际输出与期望输出不相符，则转入误差的反向传播过程。误差反向传播指输出误差的逆向逐层分摊和调节阈值使误差阶梯下降的过程。

1.3.3 支持向量机　支持向量机是一种监督学习方法，通过非线性映射，采用结构风险最小化原则，将低维空间和线性不可分的数据映射到高维空间使其成为线性可分的，再将数据在高维空间进行分类和预测[26]。支持向量机的核心是核函数，可以提高算法的速度。

2 结果与分析

2.1 干旱因子分析

降水、土壤、植被干旱因子在不同时期对干旱的表现不同。提取2002—2019年甘肃4—10月所有气象站点的VCI、TCI、PCI和VSWI指数，按月依次对4种遥感指数和1个月、3个月、6个月时间尺度的SPEI(分别由SPEI_1，SPEI_3和SPEI_6表示)进行Pearson相关性分析，进而分析单个遥感干旱指数监测农业干旱的能力以及融合多源数据的必要性，结果如表2所示。3种时间尺度的SPEI除了植被状态指数VCI在植被生长初期和末期外，其他指数的相关性均通过了0.01或0.05显著性检验。

表2 遥感指数与不同时间尺度SPEI的相关性分析

Tab. 2 Correlation analysis between remote sensing index and SPEI on different time scales

月份	SPEI_1				SPEI_3				SPEI_6			
	VCI	TCI	PCI	VSWI	VCI	TCI	PCI	VSWI	VCI	TCI	PCI	VSWI
4	0.06	0.38**	0.54**	0.30*	0.09	0.23**	0.47**	0.34**	0.09	0.28**	0.56**	0.30**
5	0.09	0.28**	0.61**	0.29*	0.07	0.24**	0.60**	0.26**	0.08	0.29**	0.53**	0.21**
6	0.12**	0.35**	0.65**	0.34**	0.10**	0.34**	0.51**	0.30*	0.21	0.33**	0.53**	0.25*
7	0.15**	0.34**	0.76**	0.40*	0.28**	0.25**	0.68**	0.29**	0.19**	0.32**	0.78**	0.32*
8	0.16**	0.24**	0.78**	0.27**	0.17**	0.27**	0.65**	0.28**	0.20**	0.25**	0.62**	0.23**
9	−0.02	0.31**	0.49*	0.24**	0.04*	0.23**	0.45*	0.29*	0.02	0.26**	0.43*	0.26**
10	0.11*	0.32**	0.55**	0.37**	0.25*	0.27**	0.50**	0.41**	0.17**	0.30**	0.56**	0.32*

注：**表示在0.01水平上显著相关，*表示在0.05水平上显著相关。VCI表示植被状态指数，TCI表示温度状态指数，PCI表示降水状态指数，VSWI表示植被供水指数。

对于降水指数PCI,其相关性在3种时间尺度的SPEI中均为最高,且在7月达到最大值,说明以降水因子建立的干旱指数在降水充足的季节其监测结果更具有参考价值。温度干旱指数TCI在植被生长初期的相关系数要高于植被生长中后期,如TCI与不同时间尺度SPEI的相关性均在6月达到最大值。在4—10月,植被干旱指数VCI与不同时间尺度SPEI的相关性均呈现先增大后减小的过程,以SPEI_3为例,在4月和5月未通过显著性检验,到7月植被生长最佳时期,相关系数达到最大,8月开始,相关性又逐渐减弱。这表明以植被因子建立的干旱指数在植被覆盖度较好的区域监测干旱能力更可靠。在月时间尺度上土壤湿度指数VSWI其相关性表现为先增大后减小再增大的趋势。其变化趋势与植被指数VCI相同,说明土壤湿度指数更适合于植被茂盛区域的旱情监测。

以上分析表明,PCI、VCI、TCI以及VSWI单个遥感干旱指数在监测农业干旱中具有局限性,虽然PCI与SPEI的相关性高,但单一降水因素不能代表干旱。因此应用先进的机器学习方法来融合多种致旱因子,构建综合监测农业干旱的模型具有重要意义。

不同时间尺度的标准化降水蒸发指数SPEI可以分析不同的干旱类型,1—6个月时间尺度的SPEI适用于气象和农业干旱,较长时间尺度的SPEI适用于水文干旱。大量研究表明3月或更长时间尺度的SPEI更能代表农业干旱,本文采用3个月时间尺度的SPEI分析农业干旱。

2.2 模型构建及验证评价

从2012—2019年甘肃气象站点中随机选取5组不同的5个站点作为模型的验证数据,其次,2002—2019年除每组被选出的5个验证站点外所有站点的SPEI_3和所在站点的VCI、TCI、PCI、VSWI、DEM、AWC和气候类型作为模型的训练数据。每组训练样本数据依次有461个、462个、465个、465个和465个。运用3种机器学习方法分别对4—10月的训练样本进行学习,每月包括5组不同的训练样本,共构建35组农业干旱监测模型。

模型构建中随机森林有两个重要参数,决策树个数n和树节点预选变量个数m;当n太小则模型欠拟合,当n足够大时模型逐渐趋于稳定。n应大于100,m的取值公式为 $\sqrt{P}$,其中P为候选特征变量,m应小于特征变量。为了选取误差最小的m,根据经验公式和不小于特征变量两个要求,逐一选择m并构建不同的回归模型。以4月的5组模型为例,当n=1000,m依次取2、3、2、2、2时,随机森林模型的误差最小,同样方法得到其余30个模型的参数。支持向量机方法构建模型时选取误差较小、分类准确率较高的径向基函数RBF,两个重要参数为核参数g和惩罚系数C。核参数g影响训练和预测速度。惩罚系数C越大,模型容易过拟合,过小容易欠拟合。经过反复试验,当支持向量机的g为0.02、惩罚系数C为10时模型是最稳定的。BP神经网络中梯度下降法是调整全局权重和阈值的关键,本文选择训练速度最快的trainlm函数,缺点是需要的内存大。

通过模型输出的模拟值与SPEI_3实测值之间的相关性分析,验证评价机器学习方法构建的农业干旱监测模型。图2为3种机器学习方法各月份第2组验证样本模拟值与实测值SPEI_3的散点图,图中随机森林、BP神经网络和支持向量机模型的模拟值分别用RF_nh、BP_nh和SVM_nh表示。由图2可以看出,4—10月份3种机器学方法模型的模拟值与SPEI_3实测值的相关系数都不小于0.89,说明3种机器学习方法构建的农业干旱监测模型都可应用于实际的旱情监测。

2.3 3种机器学习方法监测结果对比

利用决定系数R^2、均方根误差RMSE和平均绝对误差MAE来比较模型模拟结果进而分析适用于甘肃省构建模型的最佳机器学习方法。从3种机器学习方法对验证数据的预测结果统计可知(表3),在随机森林构建的35个监测模型中,随机森林模型模拟值与SPEI_3实测值的R^2在0.78~0.95,4—10月的总平均值为0.86;BP神经网络模型的R^2在0.70~0.91,4—10月的总平均值为0.82;支持向量机模型的R^2在0.73~0.91,4—10月的总平均值为0.84;同时也对每月5组模型的统计指标进行求平均值,发现随机森林模型的R^2要高于BP神经网络和支持向量机,说明随机森林模型模拟的结果对干旱指数SPEI_3的解释程度更高。随机森林、BP神经网络和支持向量机3种方法构建的模型中,RMSE和MAE的最大值分别是0.79和0.66、1.00和0.79、0.81和0.66,说明与BP神经网络和支持向量机模型相比,随机森林模型的模拟值与SPEI_3实测值的差异更

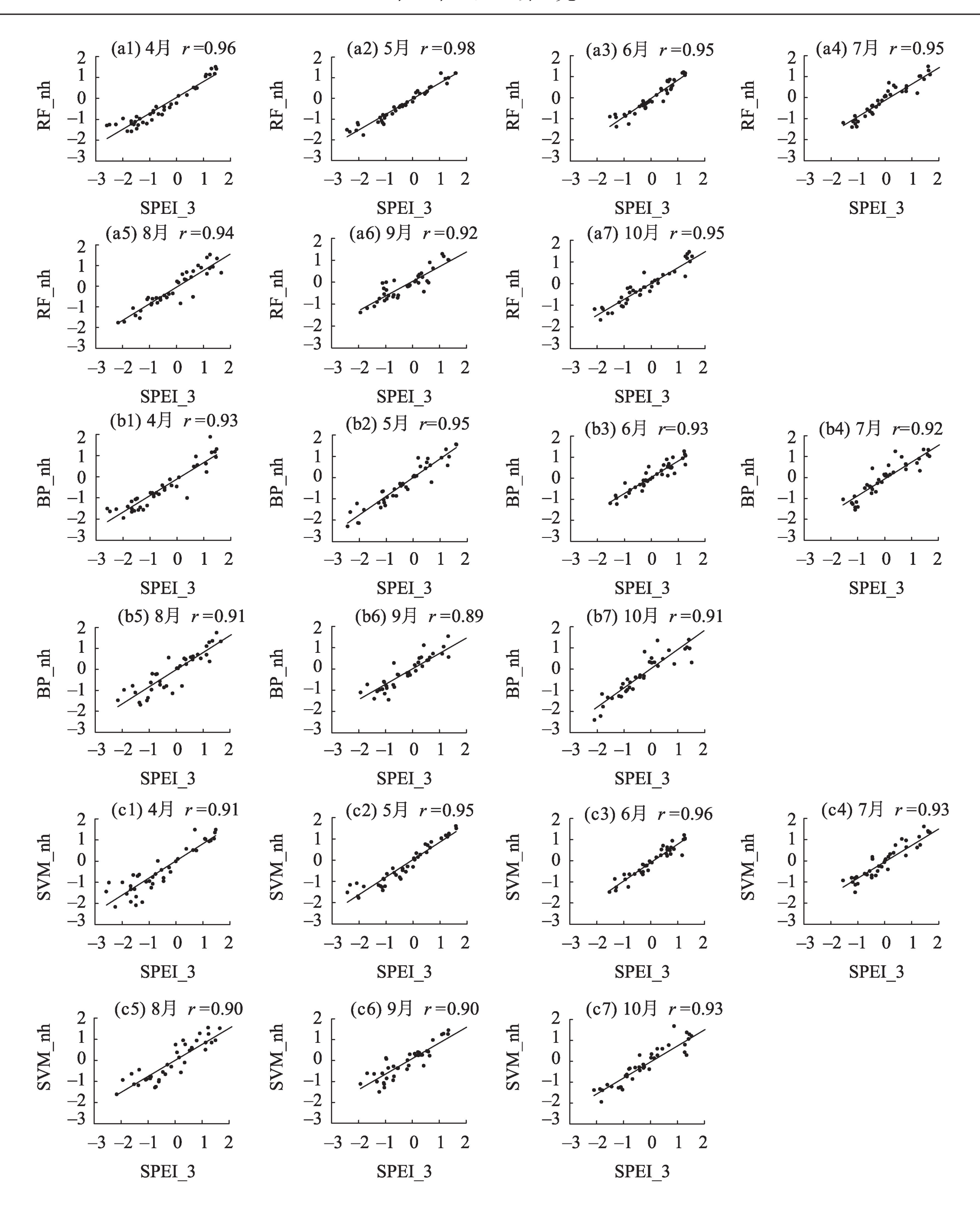


图2 模型模拟值与实测值SPEI_3散点图

Fig. 2 Scatter plot of model simulation value and measured value SPEI_3

小,表现最好,而支持向量机模型的表现则要优于BP神经网络模型。通过上述分析表明,随机森林方法构建的模型在甘肃省的农业干旱监测中表现更佳。

2.4 机器学习空间敏感性分析

甘肃气候类型复杂,随机森林、BP神经网络和支持向量机构建的3种模型受不同气候模式的影响,对干燥和潮湿地区干旱指数SPEI_3的模拟可能不同,利用K均值聚类方法根据多年降水条件将气象站点划分为干燥和湿润两种类型(表4)。首先根据两个聚类将原始数据分为聚类1(干燥)与聚类2(湿润)的训练样本和验证样本,然后分别训练BP神经网络、随机森林和支持向量机模型,最后利用决定系数R^2、均方根误差RMSE和平均绝对误差MAE来评估模型的性能(表5)。从表5可知,不管在那种环境下,随机森林模型的表现要优于其他两种模型,在干燥的环境下,BP神经网络模型的性能比支持向量机模型好,湿润环境下,与BP神经网络模型

表3 3种机器学习方法对验证数据拟合结果统计

Tab. 3 Statistics of the fitting results of the three machine learning methods on the verification data

月份	样本	随机森林			BP神经网路			支持向量机		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
4	1	0.78	0.79	0.66	0.70	1.00	0.79	0.73	0.81	0.66
	2	0.91	0.43	0.29	0.86	0.45	0.32	0.83	0.50	0.37
	3	0.84	0.50	0.39	0.80	0.58	0.42	0.85	0.49	0.39
	4	0.86	0.56	0.42	0.82	0.53	0.39	0.85	0.49	0.39
	5	0.90	0.36	0.31	0.87	0.40	0.32	0.87	0.37	0.31
	平均值	0.86	0.53	0.41	0.81	0.59	0.45	0.82	0.53	0.42
5	1	0.84	0.56	0.40	0.80	0.68	0.53	0.82	0.58	0.40
	2	0.95	0.33	0.23	0.91	0.34	0.24	0.91	0.35	0.24
	3	0.85	0.37	0.28	0.81	0.44	0.31	0.84	0.42	0.30
	4	0.88	0.36	0.28	0.81	0.47	0.37	0.84	0.38	0.27
	5	0.82	0.37	0.31	0.77	0.41	0.33	0.78	0.41	0.31
	平均值	0.87	0.40	0.30	0.82	0.47	0.36	0.84	0.43	0.30
6	1	0.80	0.38	0.29	0.77	0.39	0.30	0.76	0.40	0.32
	2	0.91	0.24	0.18	0.87	0.29	0.22	0.91	0.25	0.19
	3	0.87	0.23	0.18	0.85	0.35	0.28	0.85	0.27	0.22
	4	0.85	0.31	0.26	0.82	0.29	0.23	0.82	0.38	0.32
	5	0.82	0.36	0.29	0.82	0.35	0.27	0.79	0.35	0.28
	平均值	0.85	0.30	0.24	0.83	0.33	0.26	0.83	0.33	0.26
7	1	0.86	0.36	0.26	0.84	0.39	0.30	0.82	0.51	0.39
	2	0.90	0.38	0.27	0.85	0.41	0.31	0.87	0.39	0.29
	3	0.81	0.42	0.32	0.78	0.43	0.34	0.81	0.41	0.35
	4	0.85	0.51	0.37	0.80	0.48	0.36	0.81	0.48	0.40
	5	0.88	0.36	0.28	0.83	0.42	0.33	0.85	0.41	0.33
	平均值	0.86	0.41	0.30	0.82	0.43	0.33	0.83	0.44	0.35
8	1	0.85	0.49	0.39	0.81	0.52	0.40	0.83	0.50	0.45
	2	0.88	0.40	0.29	0.80	0.47	0.35	0.81	0.48	0.38
	3	0.90	0.34	0.29	0.87	0.38	0.31	0.88	0.36	0.29
	4	0.87	0.40	0.36	0.85	0.45	0.36	0.83	0.50	0.41
	5	0.89	0.39	0.29	0.82	0.48	0.39	0.86	0.43	0.29
	平均值	0.88	0.40	0.32	0.83	0.46	0.36	0.84	0.45	0.36
9	1	0.85	0.44	0.33	0.82	0.48	0.35	0.80	0.53	0.41
	2	0.84	0.45	0.34	0.80	0.46	0.31	0.82	0.46	0.33
	3	0.89	0.32	0.23	0.87	0.35	0.26	0.90	0.32	0.21
	4	0.89	0.34	0.28	0.82	0.43	0.32	0.86	0.42	0.31
	5	0.83	0.44	0.32	0.81	0.46	0.36	0.80	0.50	0.38
	平均值	0.86	0.40	0.30	0.82	0.44	0.32	0.84	0.45	0.33
10	1	0.84	0.44	0.32	0.76	0.54	0.44	0.83	0.46	0.35
	2	0.90	0.39	0.28	0.83	0.46	0.34	0.86	0.42	0.32
	3	0.87	0.36	0.30	0.83	0.49	0.38	0.85	0.37	0.28
	4	0.89	0.36	0.29	0.85	0.39	0.31	0.87	0.41	0.32
	5	0.86	0.37	0.26	0.82	0.42	0.33	0.84	0.41	0.31
	平均值	0.87	0.38	0.29	0.82	0.46	0.36	0.85	0.41	0.32

表4 站点聚类结果

Tab. 4 Site classification

分类	站点	年平均降水量/mm	平均温度/℃
聚类1	酒泉、张掖、武威、民勤、靖远、永昌、景泰、皋兰、高台、敦煌、瓜州、玉门镇、山丹	155.93	7.97
聚类2	崆峒、临夏、榆中、临洮、环县、西峰、岷县、武都、麦积、华家岭、会宁、天水	483.96	8.60

表5 不同空间模型性能评估

Tab. 5 Model performance evaluation in different spaces

分类	BP神经网络			随机森林			支持向量机		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
聚类1(干燥)	0.81	0.38	0.29	0.83	0.37	0.27	0.78	0.43	0.34
聚类2(湿润)	0.82	0.48	0.38	0.88	0.45	0.34	0.85	0.45	0.36

相比支持向量机模型表现更佳。此外，3种机器学习方法构建的模型在湿润环境下监测农业干旱的能力优于干燥环境下。

为了确定不同环境下模型在连续时间序列上对SPEI_3实测值模拟的性能。选择聚类1和聚类2两组验证样本中共12个站点，验证3种机器学习方法的时空差异性。图3为3种机器学习方法对2012—2019年7月的12个气象站点模拟值与实测SPEI_3的一致性分析，其中山丹、景泰、张掖、永昌、高台和皋兰为干燥环境中的站点，榆中、临洮、环县、临夏、武都和岷县为湿润环境中的站点。由图3可知，连续时间序列上RF模型在两个聚类中模拟SPEI_3实测值的表现最好。在干燥环境的站点中除张掖和高台站点外其他站点BP模型表现比支持向量机模型好，在湿润环境的站点中支持向量机模型略好，分析结果与表5相似，说明不同环境下构建的模型是可靠的。

随机森林方法可以进行变量重要性排序，因此通过该法获取两个聚类干旱因子(VCI、TCI、PCI、VSWI、AWC、气候类型和DEM)的重要性排序(图4)。气象因子PCI与TCI在干燥和湿润环境中排名均为前两名，对整个模型的影响最大，这两个气象因子的总和分别占聚类1和聚类2相对重要性的56%和55%，表明降水和温度因子是造成农业干旱的主要因素。在聚类2湿润环境中，土壤因子VSWI重要性为第3，植被因子VCI为第4。聚类1干燥环境中，植被为第3，土壤因子为第4。但受环境的影响，湿润环境下的植被长势比干燥环境下好。机器学习构建的模型在降水较多、植被覆盖度高的区域监测能力更好。

3 讨论

干旱是影响人类生活生产的自然灾害之一，精准实时的监测农业干旱是必须要解决的问题，故本文融合气象数据和遥感数据，利用随机森林、支持向量机和BP神经网络3种机器学习方法，构建了3种甘肃省综合农业干旱监测模型。已有研究表明[27]，机器学习模型可以有效的提高监测干旱的准确性，这与本文的研究是一致的。本文对模型模拟的结果与实测值SPEI_3进行比较分析，两者有很好的一致性，同时对4种单因子干旱指数与3种不同时间尺度的SPEI进行了相关性分析，模型的相关性高于单因子相关性，表明融合多源遥感数据确实提高了遥感监测农业干旱的准确性。

3种模型中，随机森林模型的模拟结果要优于支持向量机和BP神经网络模型，R^2更大，误差更小。这一结果与董婷等[28]的研究一致，随机森林模型在干旱监测中表现更优，更具有普适性。本研究中3种机器学习模型在湿润环境的表现更好，而在两种环境中随机森林模型的监测能力仍强于支持向量机和BP神经网络模型，这与Feng等[29]的研究结果是相似的，明确了本文研究的准确性和可靠性，同时也进一步表明了随机森林模型在农业干旱监测中更有潜力。但本文依然存在需要改进的地方，本文使用的降水遥感数据空间分辨率为0.25°，时间分辨率为月，都有待进一步提高。本文在干旱因子选择时没有考虑蒸发、人类活动等因素。未来需要针对这些不足进一步进行研究。

注：a1~f1为干燥环境中的气象站点，a2~f2为湿润环境中的气象站点。

图3　2012—2019年7月模型模拟值与实测SPEI_3的变化趋势

Fig. 3　The change trend of model simulation value and measured SPEI_3 from 2012 to July 2019

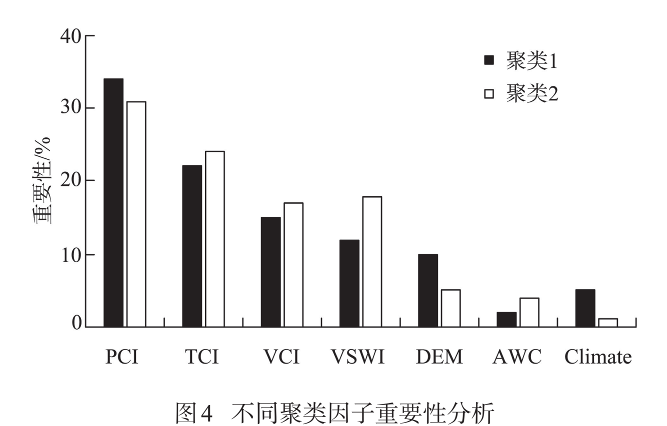


图4　不同聚类因子重要性分析

Fig. 4　Analysis of the importance of different clustering factors

4　结　论

通过采用随机森林、支持向量机和BP神经网络3种机器学习方法，集成多源数据，构建了3种甘肃省综合农业干旱监测模型，并验证了3种机器学习模型的精度，比较了3种机器学习模型在甘肃省农业干旱监测的结果以及在不同环境下模型的表现。主要得出结论如下：

（1）3种机器学习方法构建的综合农业干旱监测模型的模拟值与SPEI_3实测值的相关系数均在

0.89以上,表明构建的3种模型是有效的,均可应用于甘肃省的农业干旱监测。

(2)利用R^2、RMSE和MAE对3种模型模拟结果进行了比较,发现随机森林方法构建的模型统计指标要优于支持向量机模型和BP神经网络模型,说明RF模型能更全面可靠精准的对甘肃省的农业干旱进行监测。

(3)对比不同环境下分别构建的3种机器学习模型性能,随机森林模型在干燥和湿润环境中表现均要优于支持向量机和BP神经网络模型,说明该模型在农业干旱监测研究中更具有可靠性。结合因子相对重要性排序,发现机器学习构建的模型更适合于半干旱和植被覆盖度高的区域。

参考文献(References):

[1] 方秀琴, 郭晓萌, 袁玲, 等. 随机森林算法在全球干旱评估中的应用[J]. 地球信息科学学报, 2021, 23(6): 1040–1049. [Fang Xiuqin, Guo Xiaomeng, Yuan Ling, et al. Application of random forest algorithm in global drought assessment[J]. Journal of Geo-Information Science, 2021, 23(6): 1040–1049.]

[2] 陈少丹, 张利平, 汤柔馨, 等. 基于SPEI和TVDI的河南省干旱时空变化分析[J]. 农业工程学报, 2017, 33(24): 126–132. [Chen Shaodan, Zhang Liping, Tang Rouxin, et al. Analysis on temporal and spatial variation of drought in Henan province based on SPEI and TVDI[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(24): 126–132.]

[3] Tian L Y, Yuan S S, Quiring S M. Evaluation of six indices for monitoring agricultural drought in the south-central United States [J]. Agricultural and Forest Meteorology, 2018, 249: 107–119.

[4] 陶然, 张珂. 基于PDSI的1982—2015年我国气象干旱特征及时空变化分析[J]. 水资源保护, 2020, 36(5): 50–56. [Tao Ran, Zhang Ke. PDSI-based analysis of characteristics and spatiotemporal changes of meteorological drought in China from 1982 to 2015 [J]. Water Resources Protection, 2020, 36(5): 50–56.]

[5] Wu X, Wang P, Huo Z, et al. Crop Drought Identification Index for winter wheat based on evapotranspiration in the Huang-Huai-Hai Plain, China[J]. Agriculture Ecosystems & Environment, 2018, 263: 18–30.

[6] 张钢, 滕济端, 张立杰, 等. 综合干旱指数在崇左市的适用性分析与改进[J]. 人民珠江, 2020, 41(3): 23–29. [Zhang Gang, Teng Jiduan, Zhang Lijie, et al. Feasibility analysis and improvement of comprehensive meteorological drought index in Chongzuo[J]. Pearl River, 2020, 41(3): 23–29.]

[7] 王劲松, 郭江勇, 倾继祖. 一种K干旱指数在西北地区春旱分析中的应用[J]. 自然资源学报, 2007, 22(5): 709–717. [Wang Jinsong, Guo Jiangyong, Qing Jizu. Application of a kind of K drought index in the spring drought analysis in Northwest China [J]. Journal of Natural Resources, 2007, 22(5): 709–717.]

[8] 林慧, 王景才, 黄金柏, 等. 基于SPI和SPEI的淮河中上游流域气象干旱时空分布特征对比研究[J]. 水资源与水工程学报, 2019, 30(6): 59–67. [Lin Hui, Wang Jingcai, Huang Jinbai, et al. Comparative study on spatial and temporal distribution characteristics of meteorological drought in the upper and middle reaches of Huai river basin based on SPI and SPEI[J]. Journal of Water Resources and Water Engineering, 2019, 30(6): 59–67.]

[9] 张璐, 朱仲元, 席小康, 等. 基于SPEI的锡林河流域干旱演化特征分析[J]. 干旱区研究, 2020, 37(4): 819–829. [Zhang Lu, Zhu Zhongyuan, Xi Xiaokang, et al. Analysis of drought evolution in the Xilin River basin based on standardized precipitation evapotranspiration index[J]. Arid Zone Research, 2020, 37(4): 819–829.]

[10] 徐一丹, 任传友, 马熙达, 等. 基于SPI/SPEI指数的东北地区多时间尺度干旱变化特征对比分析[J]. 干旱区研究, 2017, 34(6): 1250–1262. [Xu Yidan, Ren Chuanyou, Ma Xida, et al. Change of drought at multiple temporal scales based on SPI/SPEI in Northeast China[J]. Arid Zone Research, 2017, 34(6): 1250–1262.]

[11] 史晓亮, 尚雨, 陈冲, 等. 淮河流域植被NDVI与干旱条件的相关性[J]. 西安科技大学学报, 2019, 39(6): 1033–1040, 1064. [Shi Xiaoliang, Shang Yu, Chen Chong, et al. Correlation of vegetation NDVI and drought conditions in Huaihe river basin[J]. Journal of Xi'an University of Science and Technology, 2019, 39(6): 1033–1040, 1064.]

[12] 李维娇, 王云鹏. 基于VCI的2003—2017年广东省干旱时空变化特征分析[J]. 华南师范大学学报(自然科学版), 2020, 52(3): 85–91. [Li Weijiao, Wang Yunpeng. An Analysis of the spatial-temporal characteristics of drought in Guangdong based on vegetation condition index from 2003 to 2017[J]. Journal of South China Normal University (Natural Science Edition), 2020, 52(3): 85–91.]

[13] 张静, 魏伟, 庞素菲, 等. 基于FY-3C和TRMM数据的西北干旱区干旱监测与评估[J]. 生态学杂志, 2020, 39(2): 690–702. [Zhang Jing, Wei Wei, Pang Sufei, et al. Monitoring and assessment of drought in arid area in Northwest China based on FY-3C and TRMM data[J]. Chinese Journal of Ecology, 2020, 39(2): 690–702.]

[14] 张红卫, 陈怀亮, 周官辉, 等. 归一化多波段干旱指数在农田干旱监测中的应用[J]. 科技导报, 2009, 27(11): 23–26. [Zhang Hongwei, Chen Huailiang, Zhou Guanhui, et al. Application of normalized multiband drought index method in cropland drought monitoring[J]. Science & Technology Review, 2009, 27(11): 23–26.]

[15] 谷佳贺, 薛华柱, 董国涛, 等. 归一化水体指数用于河南省干旱监测适用性分析[J]. 干旱地区农业研究, 2020, 38(6): 209–217. [Gu Jiahe, Xue Huazhu, Dong Guotao, et al. Applicability analysis of NDWI for drought monitoring in Henan Province[J]. Agricultural Research in Arid Areas, 2020, 38(6): 209–217.]

[16] 王立丽, 张安兵. 基于温度供水干旱指数的京津冀春旱监测及时空分布分析[J]. 测绘与空间地理信息, 2021, 44(4): 72−75. [Wang Lili, Zhang Anbing. Spring drought monitoring in Beijing-Tianjin-Hebei based on temperature water supply drought index [J]. Geomatics & Spatial Information Technology, 2021, 44(4): 72−75.]

[17] 吴志勇, 程丹丹, 何海, 等. 综合干旱指数研究进展[J]. 水资源保护, 2021, 37(1): 36−45. [Wu Zhiyong, Cheng Dandan, He Hai, et al. Research progress of composite drought index[J]. Water Resources Protection, 2021, 37(1): 36−45.]

[18] Zhang A Z, Jia G S. Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data[J]. Remote Sensing of Environment, 2013, 134: 12−23.

[19] 杜瑞麒, 张智韬, 巨娟丽, 等. 基于波文比和降水的综合干旱指数的构建及应用[J]. 节水灌溉, 2020, 45(8): 63−74. [Du Ruiqi, Zhang Zhitao, Ju Juanli, et al. Construction and application of comprehensive drought index based on Bowen ratio and precipitation [J]. Water Saving Irrigation, 2020, 45(8): 63−74.]

[20] 韩兰英, 张强, 赵红岩, 等. 甘肃省农业干旱灾害损失特征及其对气候变暖的响应[J]. 中国沙漠, 2016, 36(3): 767−776. [Han Lanying, Zhang Qiang, Zhao Hongyan, et al. The characteristics of agricultural drought disaster loss and response to climate warming in Gansu, China[J]. Journal of Desert Research, 2016, 36(3): 767−776.]

[21] Gupta S C, Larson W E. Estimating soil water retention characteristics from particle size distribution, organic mat-ter percent, and bulk density[J]. Water Resources Research, 1979, 15(6): 1633−1635.

[22] 胡鹏飞, 李净, 王丹, 等. 基于MODIS和TRMM数据的黄土高原农业干旱监测[J]. 干旱区地理, 2019, 42(1): 172−179. [Hu Pengfei, Li Jing, Wang Dan, et al. Monitoring agricultural drought in the Loess Plateau using MODIS and TRMM data[J]. Arid Land Geography, 2019, 42(1): 172−179.]

[23] 杨莎. 基于植被供水指数的锡林浩特市干旱监测分析[J]. 草原与草业, 2021, 33(3): 42−45. [Yang Sha. Drought monitoring analysis in Xilinhot City based on vegetation water supply index[J]. Grassland and Prataculture, 2021, 33(3): 42−45.]

[24] 刘冀, 张特, 魏榕, 等. 基于随机森林偏差校正的农业干旱遥感监测模型研究[J]. 农业机械学报, 2020, 51(7): 170−177. [Liu Ji, Zhang Te, Wei Rong, et al. Development of agricultural drought monitoring model using remote sensing based on bias-correcting random forest[J]. Transactions of the Chinese Society of Agricultural Machinery, 2020, 51(7): 170−177.]

[25] 李晓辉, 杨勇, 杨洪伟. 基于BP神经网络与灰色模型的干旱预测方法研究[J]. 沈阳农业大学学报, 2014, 45(2): 253−256. [Li Xiaohui, Yang Yong, Yang Hongwei. Combining BP neural network with gray model to achieve drought predicting[J]. Journal of Shenyang Agricultural University, 2014, 45(2): 253−256.]

[26] 赵国羊, 涂新军, 王天, 等. 基于人工神经网络和支持向量回归机的干旱预测[J]. 人民珠江, 2021, 42(4): 1−9. [Zhao Guoyang, Tu Xinjun, Wang Tian, et al. Drought prediction based on artificial neural network and support vector regression machine[J]. Pearl River, 2021, 42(4): 1−9.]

[27] 沈润平, 郭佳, 张婧娴, 等. 基于随机森林的遥感干旱监测模型的构建[J]. 地球信息科学学报, 2017, 19(1): 125−133. [Shen Runping, Guo Jia, Zhang Jingxian, et al. Construction of a drought monitoring model using the random forest based remote sensing[J]. Journal of Geo-Information Science, 2017, 19(1): 125−133.]

[28] 董婷, 任东, 邵攀, 等. 基于多源遥感数据和随机森林的综合旱情指标构建[J]. 农业机械学报, 2019, 50(8): 200−212. [Dong Ting, Ren Dong, Shao Pan, et al. Construction of integrated drought condition index based on multi-sensor remote sensing and random forest[J]. Transactions of the Chinese Society of Agricultural Machinery, 2019, 50(8): 200−212.]

[29] Feng P Y, Wang B, Liu D L, et al. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia[J]. Agricultural Systems, 2019, 173: 303−316.

Comparative agricultural drought monitoring based on three machine learning methods

WANG Xiaoyan, LI Jing, XING Liting
(College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, Gansu, China)

Abstract: Frequent droughts have caused serious harm to the economy and agricultural production of Gansu Province. Therefore, it is very important for the province to use advanced methods to establish accurate and reliable agricultural drought monitoring models. Three machine learning methods (random forest, BP neural network, support vector machine) are used to construct an agricultural drought monitoring model using remote sensing factors and meteorological factors. Analysis of the best model for agricultural drought monitoring in Gansu Province. At the same time, the applicability of the model constructed by machine learning in different environments was further explored. The results show that among the models constructed by the three machine learning methods, the RF model has a high coefficient of determination (0.86) and small errors (RMSE 0.40, MAE 0.31). The monitoring effect of agricultural drought is better than the model constructed by BP and SVM; The three machine learning method models constructed in the two environments, respectively, performed better in the wet environment, while the model constructed by the random forest method performed better than the other two models in monitoring drought in the two environments. The research results provide a new scientific method for agricultural drought monitoring and evaluation in Gansu Province, and are of great significance to agricultural drought research.

Keywords: agricultural drought; machine learning; SPEI; MODIS